Measurements on a Network-Studies and Techniques

Girija Limaye
07305905

April 29, 2008

Roadmap

- Motivation
- Terminologies
- Techniques
- Discussion and Examples
- Summary and Conclusion

Importance of Measurements

Some Questions:

- User: Why is the delay in mail delivery?
- Network Provider: How much time the user X is using network?
- Analyst: Which sites do people visit often?
- Web site owner: How much time it takes to open the home page after a request is given by the user?

Importance of Measurements (cntd.)

- End-user's Requirements:
  - "always ON"
  - Speed
- Network admisistrator's Objectives:
  - End-to-end performance
  - Efficient utilization of resorces
  - Diagnosis
- Network Examiner / Analyst's View:
  - Response Time, Delay, Loss, Blocking
  - Traffic pattern
  - Load
  - This analysis is useful for admins also

How to measure?

Tools

- Bandwidth: *pathchar*
- Number of hops: *traceroute*
- Traffic: *ethereal*
- Reachability: *ping*

Intuition!

- Writing Scripts
- Collecting data
- Generating graphs and tables

Terminologies

Parameter: What we measure

One Way Delay, Round Trip Time, Packet Interarrival Delay, Packet loss

Metric: What we calculate / estimate

Bandwidth (Capacity, Utilized, Available)
Traffic Pattern
Network State

Active measurements

Introducing (extra) traffic in existing network

Passive measurements

(Only) Monitoring existing network

Various Techniques

- Packet Based
  - Single Packet Probe
  - Packet Pair[1]
  - Variable packet size
  - Packet Loss
- Protocol Based
  - TTL field in IP
  - ICMP messages

[1] Those in red are covered in the report but not in the presentation

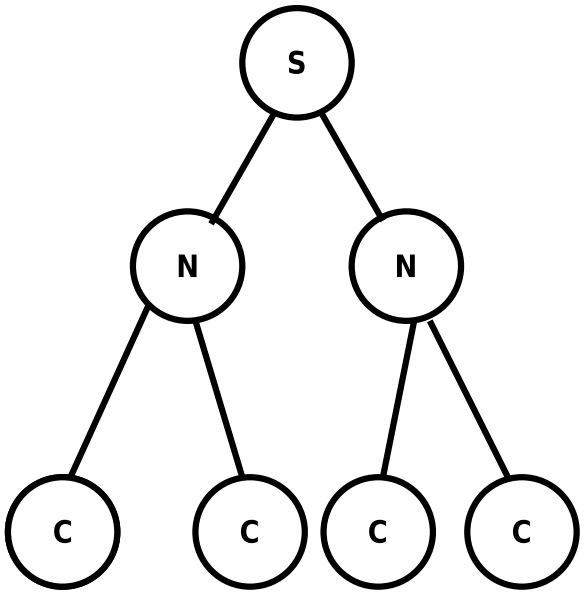
Various Techniques (cntd.)

- Delay and Delay Variance Based
  - Self Loading Packet Streams (SLoPS)
- Miscellaneous
  - Single Point Measurements inside the network
  - Statistical Modelling
  - Traces and Inferences

These techniques can be used in combination.

Bayesian Inference of Lossy link

- Problem: In such a topology, which one is a lossy link?
- Approach:
  - Links are shared
  - A path from S to C contains multiple Links
  - Loss observed at client: Function of Loss rates of these intermediate links
  - Can we model Loss rates of links in terms of loss observed at the client?



S: Sender
N: Intermediate Nodes
C: Clients

Bayesian Inference of Lossy link (cntd.)

- P_j: Loss probability observed at the client C_j
- $P_j = f(l_i)$ where $i \in$ All intermediate links from S to C_j
- Evidence P_j, conditional to the loss rates of links
- Lossy link is hidden, a Hypothesis
- Use Bayesian Inference: [2]
  $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$

In short,

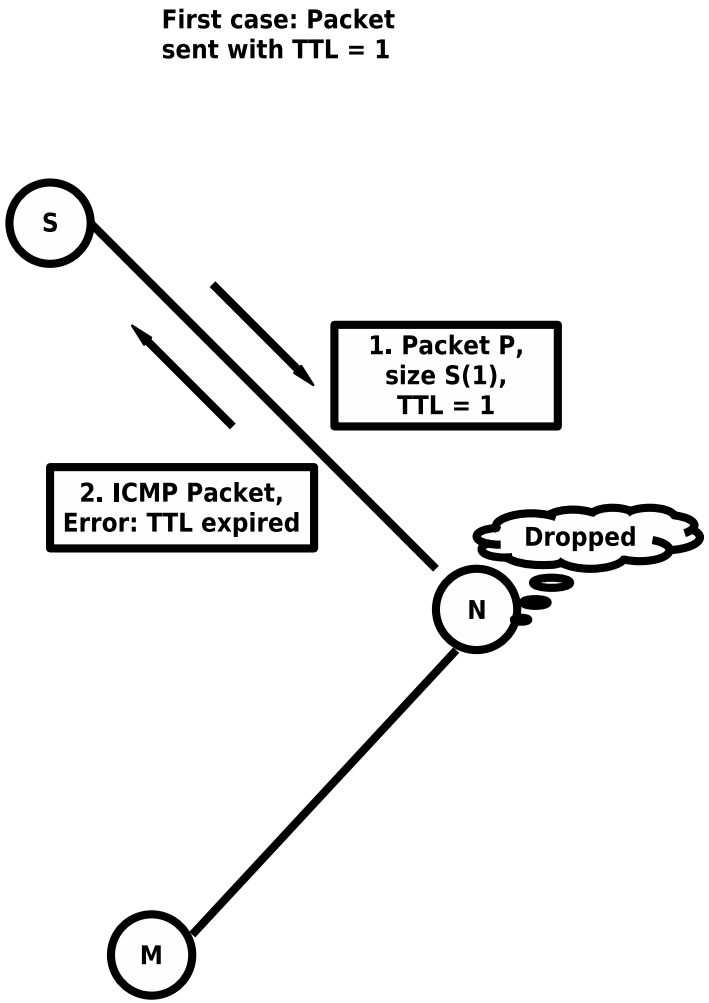
Using Packet Loss Technique, Passive Measurements
Parameter: Packet loss
Inferred: Lossy link

[2] A: Hypothesis, B: Evidence

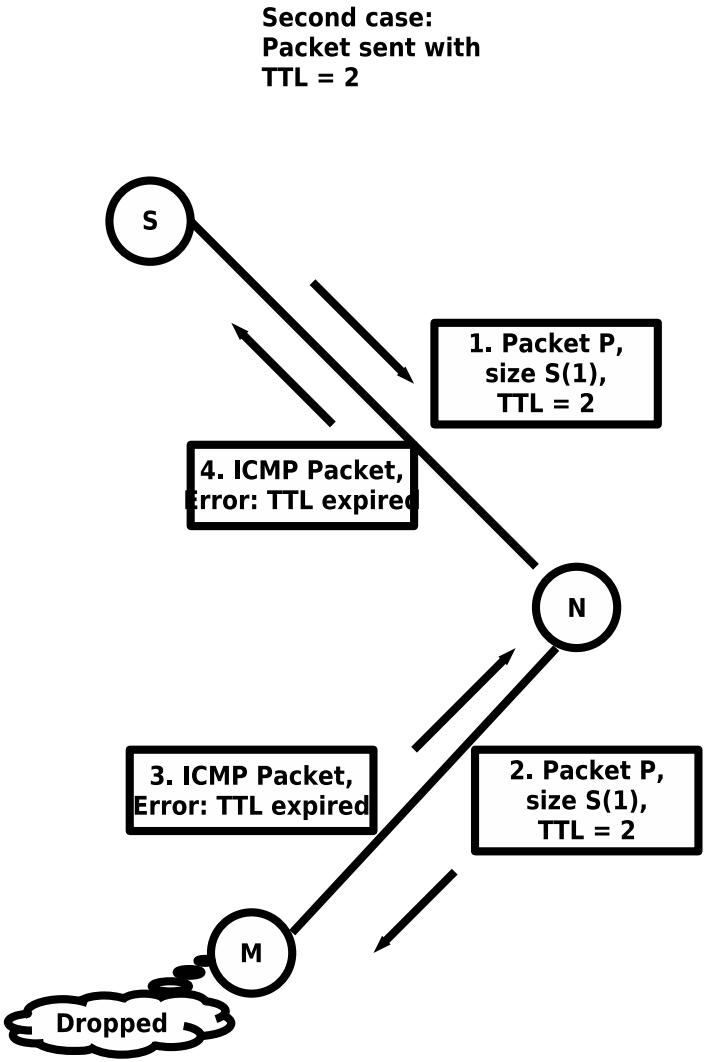
Estimate Bandwidth

- Problem: A path from S to D, multi-hop. Estimate capacity of links on the path.
- Approach:
  - Delay $\propto \frac{1}{Bandwidth}$
  - End-to-End Delay = Sum of delays over each link
  - Can we seperately find these components of delays?
  - And use them to infer Bandwidth of "a" link?

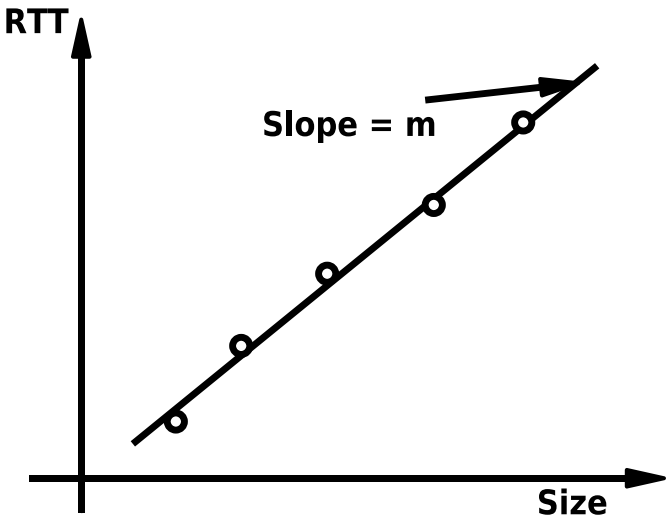
Estimate Bandwidth (cntd.)



Estimate Bandwidth (cntd.)



Estimate Bandwidth (cntd.)



RTT = delay for packet + delay for ICMP packet

So, RTT = (S / B(1)) + (S(ICMP) /B(1))

RTT = (packet size / bandwidth) + Constant

Y = mX + C

In short,

Using Single Packet Probe Technique, Active Measurements, TTL Mechanism, Varying Packet size
Parameter: RTT

Inferred: Capacity (Bandwidth)

Routing changes, Which?, How often?

- Why does end-to-end delay change?
  - Traffic Fluctuations
  - Route changed
- So, how does Routing change affect delay?
  - Number of hops increased
  - Links on new route have different properties (lower bandwidths)
- Why is it important?
  - Real Time Applications, Financial Transactions
  - VOIP, multimedia streaming, delay-sensitive

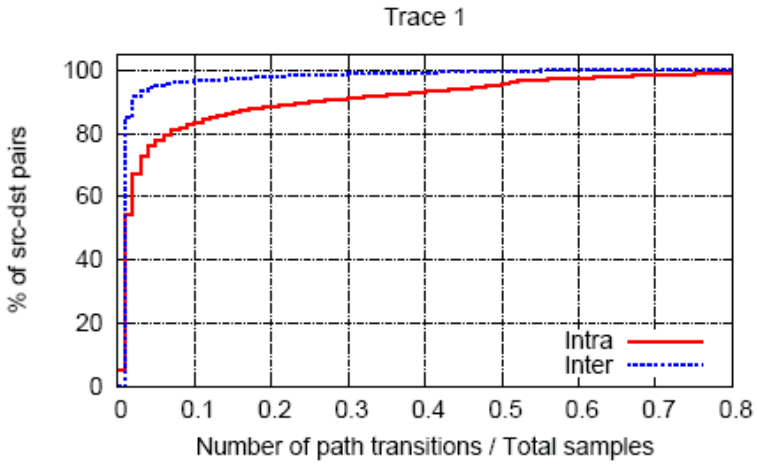
Routing changes, Which?, How often? (cntd.)

- Which Route? Interdomain / Intradomain
- Inerdomain
  - Passive Measurement
  - BGP UPDATE Messages
- Intradomain
  - Active Measurement
  - *traceroute* till convergence

Intuition

- Interdomain Routes: Have more impact on delay variation
- Long distance

Routing changes, Which?, How often? (cntd.)



Observations

- Interdomain routes change less frequently
- Interdomain routes last longer
- Intradomain Routes also have significant impact on delay variation

In short,

Combination of Active and Passive Measurements
Parameter: Delay

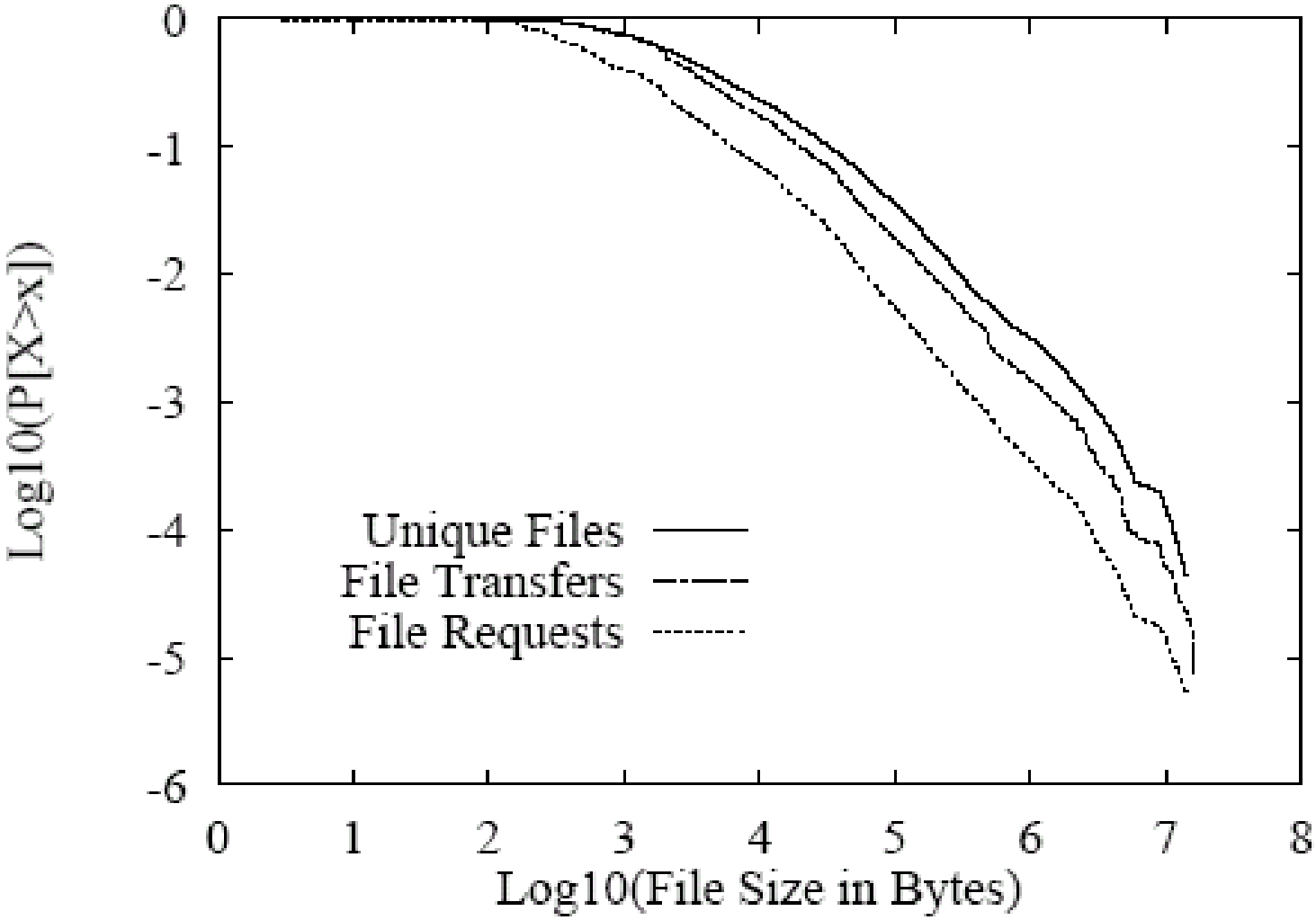
How is web traffic?

- Significance:
  - How is the traffic? TCP / UDP, Real time?
  - What is the shape, underlying distribution?
  - It is dependent on- Protocol?, User behaviour?, Scheduling?
- WWW traffic: A subset of the traffic that flows on the internet
- Heavy-tailed Distribution
  - Assumed: to be following Poisson distribution: Light-tailed
  - Heavy-tailed: Probability of values far from mean are relatively high
  - Heavy-tailed: Larger Job Sizes $\rightarrow$ High Probability

How is web traffic? (cntd.)

- Observation: Follows heavy-tailed distribution
  - Alternate periods of ON-OFF at client workstation, which are heavy-tailed
  - ON - Either receiving data / processing received data
  - OFF - No activity
  - OFF time: "Think Time" of user: Pareto $\rightarrow$ "No activity" time: Heavy-tailed
  - ON time: Available File sizes: Heavy-tailed $\rightarrow$ Files transferred: Heavy-tailed

How is web traffic? (cntd.)



How is web traffic? (cntd.)

Self-similarity (Example: Fractals)

- If for time series, m-aggregated series is taken by summing non-overlapping blocks of size m, their distribution is same as original series when original series is rescaled.
- Heavy-tailed distribution samples tend to show above property.

In short,

Passive Measurement, Data Collection
Parameter: Time, thus Delay
Inference: Traffic Pattern

Summary and Conclusion

- Need of measurements on network
- Classification: Active-Passive
- Active: Extra traffic, overhead
- Passive: Monitoring, trace collection
- Techniques
  - Protocol related: Packet pairs, ICMP packets, TTL variation
  - Statistical Modelling

Future Work

- Active measurements: Improve accuracy by reducing overhead
- Alternative Passive mechanism
- Techniques in practice
- A network measurement framework for a small campus / organization

References

- Helen J. Wang Venkata N. Padmanabhan, Lili Qiu. Server-based inference of internet link lossiness, IEEE INFOCOM 2003.
- Allen B. Downey. Using pathchar to estimate internet link characteristics, SIGMETRICS 1999.
- Mark E. Crovella and Azer Bestavros. Self-similarity in world wide web traffic: Evidence and possible causes, IEEE/ACM Transactions on Networking, 1997.

References (cntd.)

- Himabindu Pucha, Ying Zhang, Z. Morley Mao, and Y. Charlie Hu. Understanding network delay changes caused by routing events, SIGMETRICS 2007
- Wikipedia. Heavy tailed distibutions.
- Wikipedia. Self similarity.